What kind of field is AI?

What kind of field is AI?

Alan Bundy

I want to ask 'What kind of field is artificial intelligence?' and to give an answer. Why is this an important question? Because there is evidence within AI of a methodological malaise, and part of the reason for this malaise is that there is not a generally agreed answer to the question.

As an illustration, several papers in this volume draw a number of different analogies between artificial intelligence and other fields. AI is compared to physics, to chemical engineering, to thermodynamics and to psychology; in fact it is said to *be* psychology. Each of these is a very different kind of field with different kinds of methodology, criteria for assessing research and so on. Depending on which of these you think artificial intelligence is really like, you would decide what to do, how to do it, and how to assess other people's work.

1 Evidence of malaise

One of the symptoms of this malaise is a difference amongst referees of papers as to the standard which is expected for conferences, journals, etc. When I was programme chairman of a major AI conference, I noted that for more than 50 percent of the papers the referees disagreed as to whether the papers should be accepted or rejected. And this wasn't just a question of having different thresholds of acceptability, because the opinions would reverse on other kinds of papers. So clearly the referees were applying very different criteria when deciding which papers were worth accepting.

A second symptom of this malaise, which has been noted for instance by D. McDermott (1981), is the fragility of programs. Typically they run on the example in the thesis or paper, and in the words of Bobrow *et al.* (1977), on other examples they simulate total aphasia or death.

A third symptom is the poverty of published accounts. Some people have alluded to this in their criticism of papers about expert systems, where it is particularly prevalent. For instance, a paper describing an expert system for

diagnosing bee diseases, might concentrate on the nature and importance of bee diseases. At the end of the paper you might find the statement 'I built a rule based system' or 'I built a blackboard architecture', but that is all you learn about how the program actually worked. It is very difficult to build on and extend such work.

That leads me to my fourth symptom: that there is often a lack of continuity in AI research. There are in AI traditions of building on a sequence of programs, or a sequence of techniques, and these are honourable exceptions, but often one gets a one-off program which is then not continued in any way.

Part of the solution to this malaise is to identify what kind of field AI is. It is not the only answer, there are other aspects of the problem, but I don't want to dwell on these here.

2 Kinds of AI

One difficulty is that there are a number of different kinds of AI, each one of which has its own criteria and methodology. These tend to get confused together, so that different people have different conceptions of what kind of a field AI is. I want to disentangle the confusion by separating out the different kinds of AI.

The different kinds of AI correspond to different motivations for doing AI. The first kind, which has become very popular in the past five to ten years, is *applied AI*,[1] where we use existing AI for commercial techniques, military or industrial applications, i.e. to build products. Another kind of AI is to model human or animal intelligence using AI techniques. This is called *cognitive science*, or *computational psychology*. Those two kinds of AI have often been identified in the past, but there is a third kind on which I want to concentrate most of my attention. I call it *basic AI*.[2]

The aim of basic AI is to explore computational techniques which have the potential for simulating intelligent behaviour.

3 What is an AI technique?

All these definitions rely on a notion of AI technique so it is necessary for me to say something about what an AI technique is. I will be fairly catholic about this. The obvious candidate for an AI technique is an algorithm. An example would be the circumscription technique that John McCarthy has developed for non-monotonic reasoning. An older example would be means-ends analysis. The boundaries between AI algorithms and other more conventional kinds of algorithms in computer science is obviously very fuzzy. One should not expect or attempt to draw a hard and fast

boundary, but it is something to do with the potential of these techniques for simulating intelligent behaviour in a wide sense.

There are other kinds of techniques in AI, for instance, techniques for representing knowledge (e.g. situation calculus, frame systems, semantic nets). Even the new work on connectionism may be thought of as the development of new AI knowledge representation techniques.

On a wider front there are architectures for AI systems. The best example I can think of here is the blackboard architecture, but this is not a very precisely defined technique. Getting precise notions of architecture is an area of weakness in AI. There are probably some good ideas around which need tightening up and given more precise form.

Lastly, there are a collection of techniques for knowledge elicitation, for instance, protocol analysis. Notice that protocol analysis is not just a cognitive science technique; it has found application in expert systems. Knowledge elicitation techniques are also used in basic AI as a source of inspiration when developing new ideas.

4 An analogy with mathematics and physics

In order to sharpen up this three-fold division of AI I want to draw an analogy with other kinds of science. I also want to show that AI is not essentially different from other kinds of science. My analogy is drawn from mathematics and physics because they are classic sciences with which most of us are familiar, although one could draw analogies with other areas of science. In my analogy, computer science is similar to mathematics, basic AI is similar to applied mathematics and also to pure engineering, applied AI is similar to engineering, cognitive science is similar to theoretical physics, and psychology is similar to physics.

The idea is that physics and psychology are natural sciences, i.e. psychologists and physicists both try to study and model real world phenomena. Mathematics and computer science provide an armoury of techniques for building theories and/or models for these natural sciences. In applied mathematics and basic AI, people study those techniques of mathematics and computer science which are particularly useful for modelling in these natural sciences. So, in applied mathematics, one might study those partial differential equations which have proved useful for modelling hydrodynamics. In basic AI, one might study inference techniques like circumscription, frames, blackboard architectures, etc., which have been found to be useful for modelling intelligence. Then one can take those techniques and build commercial products with them. That is what is done in engineering and in applied AI.

Somewhere between the natural science of physics and applied mathematics is the area of theoretical physics, where scientists are concerned with applying the modelling techniques of applied mathematics to account for

the observations of physical phenomena. Similarly, cognitive scientists are concerned to apply the modelling techniques of basic AI to account for observations of intelligent behaviour.

It is also possible to draw an analogy between AI and engineering. Sometimes engineers are concerned with building particular houses or particular bridges or synthesizing particular chemicals, but they are also interested in developing new, general-purpose engineering techniques. I will call this activity *pure engineering*. An example is a study of new structures for building bridges without an attempt to build a particular bridge. Another example is a study of reinforced concrete in which the engineer develops various methods of doing the reinforcing and then subjects each method to a batch of tests to discover its ageing and strength characteristics. Similarly, in basic AI, researchers develop new techniques, test them and find out their interrelationships. These techniques can then be used in applied AI or in cognitive science.

The reason that it is possible to have this analogy both with applied mathematics and with pure engineering is that computer programs are strange beasts; they are both mathematical entities and artifacts. They are formal abstract objects, which can be investigated symbolically as if they were statements in some branch of mathematics. But they are also artifacts, in that they can do things, e.g. run a chemical plant. They are machines, but they are not physical machines, they are *mental machines* (Bundy, 1981). I think it is interesting that applied mathematics and pure engineering should turn out to be similar kinds of fields.

5 Is AI a science?

Whether AI is a science depends on the kind of AI and what is meant by science. We can identify two kinds of science: natural science and engineering science. In a natural science, we study some phenomena in the world and try to discover theories about them. Examples of natural sciences are physics, chemistry, biology and psychology. In an engineering science, we develop techniques and discover their properties and relations. Examples of engineering sciences are pure engineering and mathematics.

Cognitive science is a natural science. It is the study of the mind and the attempt to build theories of the mind with the aid of computational tools. Part of this study is the building of computer programs to model aspects of mental behaviour, and the comparison of the behaviour of the program with that of real minds.

Basic AI is an engineering science. It consists of the development of computational techniques and the discovery of their properties and interrelations. Part of this study is the building of computer programs that embody these techniques to: extend them, explore their properties, and generally discover more about them.

It makes sense to ask whether the theories of cognitive science are true, i.e. whether they accurately describe real minds. It does not make sense to ask whether the techniques of basic AI are true, any more than it would make sense to ask whether differentiation, group theory or internal combustion were true. It does make sense to ask whether properties of and relationships between techniques are true, e.g. whether negation as failure is a special case of circumscription, whether resolution is complete.

6 Criteria for assessing basic AI research

If you accept the analogy to physics and mathematics and the three-fold division of AI into basic AI, applied AI and cognitive science, what is the payoff? The analogy suggests criteria for assessing research in each kind of AI. It suggests how to identify what constitutes an advance in the subject and it suggests what kind of methodology AI researchers might adopt.

In demonstrating this I will concentrate on basic AI. It is also important to do this kind of exercise for applied AI and for cognitive science, but it is less urgent for these two because people have already thought about criteria for assessing research aimed at applications and psychological modelling. There has not been so much thought about criteria for assessing research in which AI techniques are developed for their own sake. In the case of applied AI there is a major criterion that what you build is in fact a successful commercial product which fills a need. In the case of cognitive science there are acid tests about doing experiments to see if the model exhibits all and only the behaviour that has been observed in humans or animals. In basic AI it is not so obvious what the criteria are.

Here are some criteria. I drew up this list by starting with particular criticisms of existing AI research and then generalizing and negating them. Although I have asserted that there is widespread disagreement about the virtues of AI research illustrated by the disagreements among referees about research papers, there is, in fact, considerable agreement about the shortcomings. People will often agree what is wrong with a piece of work without agreeing what is right. So this methodology of negating criticisms is a productive one.

The criteria discovered in this way are similar to the criteria used to assess the techniques of any engineering science. This corroboration lends support to the criteria themselves and to the classification of basic AI as an engineering science.

My first criterion is *clarity*. The technique should be describable in a precise way, independent of any program that implements it, so that it is simple to understand. Logical calculi are often useful as a language of description. They are not the only language; diagrams are another. Con-

sider the way in which minimax is usually explained: in terms of backing up scores in a game tree. Counter-examples can be found in papers that explain what a program does, but say nothing about how it does it, or in papers that describe a program only by describing how each Lisp function works. I have put clarity first because it is a necessary first step before one can go on to investigate the properties of the technique.

My second criterion is the *power* of the technique. A technique is judged to be more valuable if it has a wide range of application and if it is efficient in its operation. A typical counter-example is the Ph.D. program that works on the toy example in the thesis, but which either does not work at all or runs out of resources on other examples. If that is wrong then what is right is that the program should work on lots of examples and the more the better, and that it should work fast and use little space.

My third criterion is *parsimony*. Other things being equal, the technique which is simpler to describe should get more credit. This is often neglected in AI, but Occam's Razor ought to apply to AI techniques as much as it applies to other sciences. A counter-example would be an excessively complex program for some task where no principled attempt had been made to build a simple program for the same task.

My fourth criterion is *completeness*, i.e. that the work should be finished in some sense. For instance, a computer program should be finished and should work. We can all think of counter-examples to that. Similarly, a knowledge-representation formalism should be capable of representing the kinds of knowledge that it was intended to represent. Were it not for the counter-examples, this criterion might be thought too obvious to be worth stating.

My fifth, and last, criterion is *correctness*, i.e. that the program should behave as required, that the proofs of any theorems about the technique should be correct, etc.

7 Advances in basic AI

Given that basic AI consists of the invention and development of techniques, we can identify what constitutes an advance in the field. An obvious advance would be the invention of a new technique. Its merits can be judged according to the criteria listed in the last section. Of course, it is also an advance if a technique is improved along one of the dimensions defined by these criteria, i.e. if it is made more clear, powerful, parsimonious, complete or correct.

In assessing the merit of an advance, one must also take into account how different a new technique is from previous techniques, how surprising it is that it can be stated so clearly, parsimoniously or completely or correctly, or that it turns out to be so powerful. We can summarize this as the *novelty* of the advance.

But advances can also come from an improved understanding of the properties of, or relationships between techniques. For instance,

> a proof that a technique obeys some specification, e.g. the soundness and completeness proofs for resolution;
> the discovery of the complexity of a technique;
> a demonstration that one technique is a special case of another technique;
> a demonstration that what appears to be one technique in the literature, turns out, under closer examination, to be two or three slightly different techniques which deserve different names;
> or, vice versa, that two or more apparently different techniques are essentially the same;
> a demonstration that a technique applies to a new domain;
> an exploration of the behaviour of a technique on a range of standard examples.

Again we want to use the novelty of the advance as one measure of its merit.

8 A methodology for basic AI

What methodology does this suggest for basic AI? To answer this question we have to take a partly descriptive and a partly normative stance, that is, we need to see what methodology is actually followed in AI and then see what modifications to this methodology are suggested by our analysis. The major methodology in AI is *exploratory programming*; one chooses some task which has not been modelled before, and then, by a process of trial and error programming, develops a program for doing this task. Suppose you want to build a program which can suggest new Chinese recipes, then you might set to work developing such a program. You will dream up a scenario of how a particular recipe might be generated, implement the procedure this suggests, and then test it on further examples. This testing will show up inadequacies and suggest ways of modifying and improving the program. This trial and error cycle will continue until you are satisfied with the program.

At this stage there ought to be an *analysis* of the program in order to try to tease out in a more precise way the techniques which underlie it. Suppose that the Chinese recipe program works by matching some set of ingredients against an existing list of recipes and trying to get a close match. Somewhere in the program there must be some kind of analogical matching routine. Careful analysis of the program will reveal what form it takes.

It is quite likely that this analogical matching technique will be *ad hoc* and domain specific. So the next step is to *generalize* and to formulate a *rationally reconstructed* algorithm which does the same thing but is more

robust. We then need to explore the properties of this rationally reconstructed technique according to the list given in the last section, i.e. construct a specification and verify the technique, discover its complexity, explore its relationship to other techniques, apply it to new domains and test it on standard examples. This exploration will suggest further generalizations and improvements.

Analysis should also help identify the weaknesses of existing techniques and hence suggest what problems we should focus on in the future. Where the weaknesses cannot readily be met by the extension of existing techniques then they might be addressed by exploratory programming. A task must be identified for which the existing techniques are inadequate because of these weaknesses, and where there is some prospect that progress might be made. An attempt should then be made to build a program for this task. Thus the methodology cycles.

These latter steps in the methodology of AI – of analysis, generalization, rational reconstruction and exploration of properties – are not generally recognized or practised. It follows from my analysis of basic AI that they are very important and that they deserve a lot more attention.

9 Conclusion

To sum up: the malaise which I have identified in AI can be cured, in part, by careful definition and sub-division of the field into basic AI, applied AI and cognitive science. For each kind of AI we can identify different sets of criteria for assessing research, different notions of what constitutes an advance and different methodologies for conducting research. I have identified these for basic AI. If we do not subdivide the field then these criteria and methodologies become entangled, leading to a confusion in the way that AI research is conducted and judged.

This is not to suggest that there are not researchers with multiple motives who would like to contribute to more than one kind of AI, and that there are not pieces of work which do contribute to more than one kind of AI. Judging such work is difficult, but it can be done by separating out the work's contributions to basic AI, applied AI and cognitive science.

Notes

1 Elsewhere (Bundy, 1983) I have called it *technological AI*.
2 Elsewhere (Bundy, 1983) I have called it *mainstream AI*.

Programs in the search for intelligent machines: the mistaken foundations of AI

Eric Dietrich

Of course, unless one has a theory, one cannot expect much help from a computer (unless *it* has a theory). *Marvin Minsky*

1 Introduction

Computer programs play no single role in artificial intelligence. To some, programs are an end; to others, they are a means. These two groups might be thought to contain those who think AI is an engineering discipline, and those who think AI is a science. This is only partially true; the real situation is more complicated.

The first group is by far the largest and contains many of the most prominent AI researchers. For example, in his book *Problem Solving Methods in Artificial Intelligence*, Nils Nilsson states that

> Future progress in [artificial intelligence] will depend on the development of both practical and theoretical knowledge ... As regards theoretical knowledge, some have sought a unified theory of artificial intelligence. My view is that artificial intelligence is (or soon will be) an engineering discipline since its primary goal is to *build* things. (*1971, pp. vii-viii, his emphasis*)

Barr and Feigenbaum (taking a slightly more cautious position) also claim that "whether or not [AI leads] to a better understanding of the mind, there is every evidence that [AI] will lead to a new *intelligent technology*" (1981/1982, p. 3, their emphasis).

Many researchers who see themselves as theorists or scientists also belong in this group because they think that the ultimate goal of their work on theory is to produce a computer program that does something useful, whereas in other disciplines, the goal of theorists and scientists is to produce a *theory*. Roger Schank, for example, in the preface to *Conceptual Information Processing* (1975a) states